AD-A070 205 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
ASYMPTOTIC DISTRIBUTIONS OF SLOPE OF GREATEST CONVEX MINORANT E--ETC(U)
APR 79 S LEURGANS DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1946 NL

1 OF 1
AD
A070205

END
DATE
FILMED
7-79
DDC

(12) LEVEL II

AD A070205

MRC Technical Summary Report # 1946

ASYMPTOTIC DISTRIBUTIONS OF SLOPE OF GREATEST CONVEX MINORANT ESTIMATORS

Sue Leurgans

See 1473

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

April 1979

(Received March 7, 1979)

DDC
RECEIVED
JUN 21 1979
B

DDC FILE COPY

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, D. C. 20550

79 06 20 041

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ASYMPTOTIC DISTRIBUTIONS OF SLOPE OF GREATEST CONVEX MINORANT ESTIMATORS

Sue Leurgans

Technical Summary Report # 1946

April 1979

ABSTRACT

Isotonic estimation involves the estimator of a function which is known to be increasing with respect to a specified partial order. For the case of a linear order, a general theorem is given which simplifies and extends the techniques of Prakasa Rao (1966) and Brunk (1970). Sufficient conditions for a specified limit distribution to obtain are expressed in terms of a local condition and a global condition. The theorem is applied to several examples. The first example is estimation of a monotone function μ on $[0,1]$ based on observations $(i/n, X_{ni})$, where $EX_{ni} = \mu(i/n)$. In the second example, i/n is replaced by random T_{ni}. Robust estimators for this problem are described. Estimation of a monotone density function is also discussed. It is shown that the rate of convergence depends on the order of first non-zero derivative and that this result can obtain even if the function is not monotone over its entire domain.

AMS (MOS) Subject Classifications: 60F05, 62E20, 62G05, 62G20.

Key Words: Isotonic estimation, asymptotic distribution theory.

Work Unit Number 4 - Probability, Statistics, and Combinatorics.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024. This material is based upon work supported by the National Science Foundation under Grant Nos. MCS77-16974 and MCS78-09525.

79 06 20 041

SIGNIFICANCE AND EXPLANATION

In many experiments one would expect that an increase in "input" will produce an increase in "output". For instance, the greater the vitamin concentration, the faster the growth of organisms; the greater the force applied to a rod, the greater the elongation. However, due to random effects, experimental results may not show the expected monotonic behavior.

The most common method for dealing with this situation is to use curve-fitting (e.g., by least-squares), assuming some parametric form (e.g., polynomial behavior). However there are many situations where this is not appropriate.

This paper discusses the estimation of functions which are known to be monotone, but which are not assumed to have any particular parametric form. The exact distributions of these estimators is very complicated, but some limiting distributions are known. The paper proves a relatively abstract theorem, which can then be used to obtain all the known limiting distributions, and to extend these results. Some new estimators are also described. The theorem is applied to estimation of monotone functions and to estimation of monotone density functions.

The results are applied to one of the most common non-parametric methods for dealing with situations where failure of the data to exhibit the expected monotonicity is regarded as a sampling artifact, namely the isotonized mean: two neighbouring observations that do not have the expected monotonic behavior are replaced by two numbers, each equal to the mean of the observations. This procedure is repeated until the estimator is a monotone function. The results in this paper show that the isotonized mean is sensitive to extreme values. It is shown the linear combinations of order statistics can be used to obtain estimators which are more robust than the isotonized mean.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ASYMPTOTIC DISTRIBUTIONS OF SLOPE OF GREATEST
CONVEX MINORANT ESTIMATORS

Sue Leurgans

1. Introduction

Suppose for each of n independent variables X_i there is a known t_i such that the distribution of X_i is believed to be determined by and to vary with t_i. Let F_{t_i} denote the cumulative distribution function (CDF) of X_i. Let $\theta(\cdot)$ be a specified functional on a subspace of cumulative distribution functions. θ induces μ, a real-valued function on the space of t's by $\mu(t) = \theta(F_t)$. μ is an isotonic function if there is a partial order on the space of t's such that whenever t is "greater than or equal to" s, $\mu(t) \geq \mu(s)$. This paper concentrates on the case in which the t_i are real numbers with the usual ordering and μ an isotonic function is equivalent to μ a non-decreasing function. An isotonic (or monotone) estimator of μ will be an estimator which always has the known monotonicity, but is not restricted to a particular functional form. Use of an isotonic estimator is appropriate if the order relation is certain, that is, if the failure of the observations to exhibit the specified order is an artifact of the randomness of the observations dominating the unknown underlying deterministic increasing function.

The least-squares solution to this problem has been known for some time. Ayer, et.al. (1955) and van Eeden (1956) describe an estimator $\hat{\mu}_n(t)$ (the isotonized mean) which is the monotone function with smallest error sum of squares $\left(\sum_{i=1}^{n}(X_i - \mu(t_i))^2, \mu \text{ nondecreasing}\right)$. $\hat{\mu}_n$ adaptively pools observations until the group means are increasing. Barlow, et.al. (1972, Chapter 1) discuss several algorithms for computation of this estimator. We shall use the fact that $\hat{\mu}_n(s)$ is the left hand slope of the greatest convex minorant of the cumulative sum process of the X's $(\{(j, \sum_{i=1}^{j} X_i), 0 \leq j \leq n\})$. The asymptotic distribution of this estimator was stated by Brunk (1970). This paper shows that Brunk's result can be sharpened and extended through the use of a theorem on the distribution of the slopes of greatest convex minorants

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024. This material is based upon work supported by the National Science Foundation under Grant Nos. MCS77-16974 and MCS78-09525.

Section
Section
DISTRIBUTION/AVAILABILITY CODES
Dist SPECIAL
A

of processes. This theorem can also be used to extend the results of Prakasa Rao (1969) on estimation of monotone densities, as well as to obtain asymptotic distributions of other estimators. The general theorem is stated in section 2, applications of the theorem are indicated in section 3, and the general theorem is proved in section 4. The final section contains a discussion of the relationship of the results described here to other research.

2. The General Theorem

The asymptotic distribution of these estimators is of interest because the finite sample distributions are especially complicated in all but the very simplest cases. However, to obtain limiting results, it is necessary to specify how the limits are obtained. If, for example, the set of t's to which X's correspond is fixed (and hence finite), while the number of X's observed at each t becomes infinite, and if the mean of those X's corresponding to a particular t converges to $\mu(t)$ and μ takes on a distinct value at each of the t's for which observations are recorded, then these means are asymptotically consistent, asymptotically independent, and asymptotically normal (if rescaled in the usual manner) (See Parsons (1975), for further discussion of this case).

This paper concentrates on the case in which the number of distinct t's at which observations are made becomes infinite. Exact conditions on the t's appear below. Meanwhile we assume that for each n we observe $\{(T_{ni}, X_{ni}), i = 1(1)n\}$ where $X_{ni} \sim F_{T_{ni}}$. We can assume that the observations are indexed so that T_{ni} increases (strictly) with i for every n. If the T's are random, X_{ni} is thus the concomitant of the ith order statistic of the T's. Since the isotonicity of the underlying function μ is preserved by monotone increasing functions of T, it will often be convenient to work with $\{(t_{ni}, X_{ni})\ i = 1(1)n\}$, where $t_{ni} = i/n$. The t_{ni} are essentially equally spaced in the unit interval. If F_n is the right continuous (empirical) distribution function of the T_{ni}, and Q_n is the left continuous quantile function, $t_{ni} = F_n(T_{ni})$ and $T_{ni} = Q_n(t_{ni})$.

As suggested by the isotonized mean, we may wish to work with estimators of the form $\mu_n(s) = \text{slogcom}(s)\ \{(t, Z_n(t))\ t \in T\}$ where $\text{slogcom}(s)\{A\}$ is the left-hand slope at s of the greatest convex minorant of the set of points A, $Z_n(t)$ is a random continuous process and T is an interval containing s. Theorem 1 states that if the process Z_n satisfies two conditions, then the asymptotic behavior of $\mu_n(s)$ is known. While the conditions look quite complicated, they can be described intuitively and verified in practice. Before examining the conditions, it is useful to note that the proof uses the approximate estimators $\mu_{nc}(s) = \text{slogcom}(s)\ \{(t, Z_n(t)), |t-s| \le 2cn^{-p}\}$. $\mu_{nc}(s)$ is seen to be a local version of $\mu_n(s)$.

The first condition Z_n is that the increments of Z_n stay above certain lines over certain regions with sufficiently high probability. Remark that these lines depend on n and c, although this dependence is suppressed in some of the notation. Therefore weak convergence of the Z_n processes will <u>not</u> imply Condition 1.

<u>Condition 1</u> (Hitting Times)

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} P\{Z_n(t) - Z_n(s) < L_i(t), \text{ some } t \in I_i\} = 0 \quad i = 1(1)4$$

where $L_i(t)$ is a line and I_i an interval.

$$I_1(t) = (-\infty, s] \cap T$$

$$I_2(t) = [s + 2cn^{-p}, \infty) \cap T$$

$$I_3(t) = [s, \infty) \cap T$$

$$I_4(t) = (-\infty, s - 2cn^{-p}] \cap T .$$

The first two lines are defined as follows:

$$L_1(t) = -t(n,c) - (s-t)\bar{\mu}(n,c)$$

$$L_2(t) = -t(n,c) - (s_n - t)\bar{\mu}(n,c),$$

where

$$\bar{\mu}(n,c) = \mu(s) + (2cn^{-p})^{1-p/2p} \rho(s)$$

$$s_n = s + 2cn^{-p}$$

$$t(n,c) = \zeta(1 - 2^{(p-1)/2p})2^{(1-p)/2p} \rho(s)\, n^{-(p+1)/2} c^{(1+p)/2p}, \quad 0 < \zeta < 1$$

for some constants $\mu(s)$, p, $\rho(s)$ and ζ.

L_3 is obtained by using the formula for L_1 with $\bar{\mu}(n,c)$ is replaced by $\mu(s) - (2cn^{-p})^{1-p/2p}\rho(s)$.

L_4 is obtained from L_2 by making the same substitution and replacing s_n by $s - 2cn^{-p}$.

The second condition is that a suitably renormalized version of Z_n converge to a Wiener process about a convex function. This condition will be used to obtain the limiting behavior of $\mu_{nc}(s)$ for c fixed and n large. Thus the global first condition will be used to show that the local behavior of Z_n determines the asymptotic behavior of μ_n.

Condition 2 (Local Weak Convergence)

$$\left(t,\ \frac{n(Z_n(s + 2cn^{-p}t) - Z_n(s)) - 2cn^{1-p}t\ \mu(s)}{(2cn^{1-p})^{1/2}\ \sigma(s)}\right)_{|t| \le 1} \xrightarrow[n\to\infty]{w}$$

$$\left(t,\ W(t) + \frac{\rho(s)}{\sigma(s)}\ (2c)^{1/2p}|t|^{(1+p)/(2p)}\right)_{|t| \le 1}$$

where p, $\mu(s)$, $\rho(s)$ and $\sigma(s)$ are constants ($\rho(s)$ and $\sigma(s)$ positive); $W(t)$ is a two-sided standard Wiener process on $[-1,+1]$; and the convergence is weak convergence on $C[-1,+1]$.

Theorem 2.1

If for some constant $\mu(s)$ and positive constants $\sigma(s)$, $\rho(s)$ and p, the processes Z_n satisfy Conditions 1 and 2 above and $\mu_n(s) = \text{slogcom}(s)\ \{(t, Z_n(t)),\ t \in T\}$, then

(1) $$\frac{n^{\frac{1-p}{2}}(\mu_n(s) - \mu(s))}{(\sigma(s))^{1-p}(\rho(s))^{p}} \xrightarrow{d} X^{(p)} ,$$

where $X^{(p)} \overset{d}{=} \text{slogcom}(0)\ \{(t,\ W(t) + |t|^{\frac{1+p}{2p}}\quad |t| < \infty\}$ and $W(t)$ is a standard Wiener process on $\mathbb{R}$ with $W(0) = 0$.

The following extension of Theorem 2.1 will be used for Corollary 3.2 below. It will not be proved explicitly, but follows from a routine modification of Step 3 of the proof.

Corollary 2.2

Under the conditions of Theorem 2.1, if $D_n = o_p(n^{-p})$ and $\tilde{\mu}_n(s) = \mu_n(s+D_n)$, then the conclusion of Theorem 2.1 holds for $\tilde{\mu}_n$.

We shall see below that the case $p = 1/3$ is most common in applications. In this case, the distribution of $X^{(p)}$ can be described without use of convex minorants. As stated by Prakasa Rao (1969), the distribution of $X^{(1/3)}$ is that of $T/2$, where T is the random value at which $W(t) - t^2$ attains its maximum. Chernoff (1964, Theorem 1, p. 37) proves that T has a density of the form $h(x)h(-x)$, where h is a function involving partial derivatives of a particular solution of the heat equation.

3. Applications of the Theorem

This section consists of four examples of the application of Theorem 1. We shall see that Theorem 1 can be used to reduce the derivation of the asymptotic distribution of an estimator of a monotone function to the verification of specific conditions, each of which is suited to more fundamental probabilistic approaches.

Example 1: The Isotonized Mean, Equally spaced deterministic t's.

Let the functional $\theta(\cdot)$ operate on the space of cumulative distribution functions with finite expectations by assigning to each CDF its expectation. The induced function μ satisfies $\mu(t) = EX_t$. The isotonized mean $\hat{\mu}_n$ alluded to in section 1 is a natural estimator. Since $\hat{\mu}_n(s) = \text{slogcom}(s)\ \{(j, \sum_{i=1}^{j} X_i),\ 0 \le j \le n\} = \text{slogcom}(s)\ \{(j/n, \sum_{i=1}^{j} X_{ni}/n),\ 0 \le j/n \le 1\}$, define $Z_n(t)$ to be the random function defined by linear interpolation between $\{(t_{nj}, \sum_{i=1}^{j} X_{ni}/n)\ 0 \le j \le n\}$. (Recall $t_{nj} = j/n$.) Thus $Z_n(t)$ is a normalized cumulative sum process. Theorem 1 will be applied to give the following result:

Corollary 1

Assume the following six conditions are met:

1. X_{ni}, $i = 1(1)n$ are mutually independent, for each n.
2. $EX_{ni} = \mu(i/n)$.
3. $\text{Var}\ X_{ni} = \sigma^2 < \infty$ (and $\sigma^2 > 0$) and $(X_{ni} - \mu(i/n))^2$, $1 \le i \le n$, $n \ge 1$ are uniformly integrable.
4. $0 < s < 1$ such that $\sup_{t \le s-\delta} \mu(t) < \mu(s) < \inf_{t \ge (s+\delta)} \mu(t)$ for some $\delta > 0$, and μ is increasing in a neighborhood of s.
5. μ has an Nth order derivative at s.
6. N is the smallest positive (finite) integer with $\mu^{(N)}(s) > 0$.

Then

$$n^{N/(2N+1)} \left[\frac{(N+1)!}{\mu^{(N)}(s)\sigma^{2N}}\right]^{1/(2N+1)} (\hat{\mu}_n(s) - \mu(s)) \xrightarrow[n\to\infty]{d} X^{((2N+1)^{-1})} .$$

Proof: The conditions of Theorem 1 are checked with $p = (2N+1)^{-1}$, $\sigma(s) = \sigma$, $\rho(s) = \mu^{(N)}(s)/((N+1)!)$, and Z_n the normalized cumulative sum process. Denote $\bar{\mu}(n,c) - \mu(s)$ by $\varepsilon(n,c)$. For notational convenience, we ignore the negligible effect of ns failing to be an integer.

We sketch the verification of the first Hitting Time Condition. (The others are routine variations.) In this example, the first Hitting Time Condition reduces to

$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} P\{\sum_{i=k}^{ns} X_{ni}/n < -t(n,c) - (s-k/n)\bar{\mu}(n,c),$ some $k \leq ns\}$, which involves $q(n,c) =$ $P\{\sum_{i=k}^{ns} (X_{ni}-\mu(t_{ni})) > nt(n,c) + (ns-k)\varepsilon(n,c) + \sum_{i=k}^{ns} (\mu(t_{ni})-\mu(s)),$ some $k \leq ns\}$, the probability that a cumulative sum process crosses a line, where the sequence of cumulative sums depends on n and the line depends on n and on c. Since the fourth assumption of the corollary implies that $\mu(t_{ni}) = E(X_{ni}) \leq \mu(s)$ for $i \leq ns$, $q(n,c) \leq$ $P\{\sum_{i=k}^{ns} (X_{ni}-\mu(t_{ni})) > nt(n,c) + (ns-k)\varepsilon(n,c),$ some $n \leq ns\}$. This last expression can be written as $P\{S_{n\ell} > nt(n,c) + \ell\varepsilon(n,c),$ some $0 \leq \ell \leq ns\}$, where $S_{n\ell}$ is the ℓth cumulative sum of n independent random variables with variance σ^2. Using the Dubins-Savage inequality (see Dubins & Savage (1965) or Dubins & Freedman (1965)) or the Hájek-Rényi Inequality applied to the submartingales $S_{n\ell}^2$, $0 \leq \ell \leq n\}$ with constants $c_k = \varepsilon(n,c)/(\sigma^2 + nt(n,c)k)$ (see Chow, Robbins, and Siegmund (1970), p. 25), it can be shown that $P\{S_{n\ell} > nt(n,c) + \ell\varepsilon(n,c),$ some $0 \leq \ell \leq ns\} \leq (1 + \varepsilon(n,c)nt(n,c)/\sigma^2)^{-1} = (1 + c^{2N+1}2^N(2^N-1)\rho^2(s)\zeta/\sigma^2)^{-1}$. This implies $\lim_{c\to\infty} \overline{\lim_{n\to\infty}}\, q(n,c) = 0$, as desired.

After suitable changes of notation, the weak convergence condition is seen to require the weak convergence of

$$(2) \qquad \left(\left(\frac{\ell}{k(n)}\right), \sum_{i=1}^{\ell} \frac{(X_{n,ns+i} - \mu(\frac{ns+i}{n}))}{(k(n))^{1/2}\sigma} + \sum_{i=1}^{\ell} \frac{(\mu(s+i/n) - \mu(s))}{\sigma(k(n))^{1/2}}\right)_{0\leq|\ell|\leq k(n)} ,$$

where $k(n) = 2cn^{1-p}$. Assumption 3 ensures that the Lindeberg Condition holds for $\{(X_{n,ns+i} - \mu((ns+i)/n)),\ 1 \leq i \leq k(n)\}$, which in turn implies that the first component of (2) converges weakly to Brownian motion (see Billingsley (1968), p. 77, problem 10.1). The second component is deterministic and converges uniformly to $\rho(s)t^{N+1}(2c)^{N+1/2}/\sigma$, where $\rho(s) = \mu^{(N)}(s)/(N+1)!$. Therefore the local weak convergence condition is satisfied, and the proof of Corollary 1 is complete.

□

Example 2: Isotonized Mean; Random t's

Let $\theta(\cdot)$ and μ be as in Example 1. Corollary 2 shows that the assumption of equally spaced deterministic t's can be relaxed. In what follows, F_n will be the usual right-continuous function and Q_n will be the left-continuous empirical quantile function. Note that $t_{ni} = F_n(T_{ni})$ and $T_{ni} = Q_n(t_{ni})$ if there are no tied T_{ni}'s.

Corollary 2.

Assume that conditions 3 through 6 of Corollary 1 and the following three conditions are met:

1. $\{T_{ni},\ 1 \leq i \leq n\}$ are the order statistics of a sample of size n from a distribution F which possesses a positive derivative $f(s)$ at s.
2. $E(X_{ni}) = \mu(T_{ni})$ and $\{X_{ni} - \mu(T_{ni}),\ 1 \leq i \leq n\}$ is a set of mutually independent random variables for each n.
3. For every n, $\{(X_{ni} - \mu(T_{ni})),\ 1 \leq i \leq n\}$ and $\{T_{ni},\ 1 \leq i \leq n\}$ are independent sets of random variables.

Then

$$n^{N/(2N+1)} \left[\frac{(N+1)!f(s)}{\mu^{(N)}(s)\sigma^{2N}}\right]^{1/(2N+1)} (\tilde{\mu}_n(s) - \mu(s)) \rightarrow X^{((2N+1)^{-1})}$$

where $\tilde{\mu}_n(s)$ is the isotonized mean based on $\{(T_{ni}, X_{ni}),\ 1 \leq i \leq n,\ n \geq 1\}$ evaluated at s.

Note that the limiting distribution is exactly the same as in Corollary 1, except for the presence of $f(s)$ in the normalizing constant.

Proof: Let p, $\sigma(s)$, and Z_n be as in the proof of Corollary 1, and let $\rho(s) = \mu^{(N)}(s)/(f(s)(N+1)!)$. Recall that $\tilde{\mu}_n(s) = \mu_n(F_n(s))$, where μ_n is the isotonized mean based on $\{(t_{ni}, X_{ni}), 1 \leq i \leq n\}$. Note that $d_n = F_n(s) = \theta_p(n^{-1/2}) = o_p(n^{-p})$, since $p < 1/2$. Since the conditions can be verified for s or for $s+D_n$, we choose the more convenient form for each condition. By substitution, the Hitting Time Condition involves probabilities $q(n,c)$ which are obtained from those of the first corollary by substituting T_{ni} for t_{ni} and $nF_n(s)$ for ns. Since the T_{ni}'s are increasing in i, $EX_{ni} = \mu(T_{n,i}) < \mu(T_{n,nF_n(s)+1}) = \mu(Q_n(F_n(s)+1/n)) < \mu(s)$ for $i \leq nF_n(s)$, and Hitting Time Condition follows from the argument for Corollary 1.

The weak convergence condition involves the convergence of the process defined by linear interpolation between the points of

$$(3) \qquad \left(\frac{\ell}{k(n)}, \sum_{i=1}^{\ell} \frac{(X_{n,nF(s)+i} - \mu(T_{n,nF(s)+i}))}{(k(n))^{1/2}\sigma} + \sum_{i=1}^{\ell} \frac{(\mu(T_{n,nF(s)+i}) - \mu(s))}{\sigma(k(n))^{1/2}}\right)_{0 \leq \ell \leq k(n)} .$$

The first component converges, as before, and is independent of the second component, which is now a random process. The conditions of the corollary imply that the second component converges weakly to the nonrandom process $(\mu^{(N)}(s)/(f(s)(N+1)!)t^{N+1}/\sigma$. A proof of this fact can be based on the observation that since μ is monotone, $\{\mu(T_{nj}), 1 \leq j \leq n\}$ is the set of order statistics of a sample from the CDF $F \circ \mu^{-1}$. Therefore the second component requires a local weak convergence of cumulative sums of spacings. If $F \circ \mu^{-1}$ is an exponential distribution the fact that the spacings are independent exponentials can be used to construct an embedding in a Brownian motion from which the result follows. The differentiability of F at s is used for a Taylor series expansion.

Example 3: Smoothly weighted linear combinations of order statistics; equally spaced observations

Let J be a smooth (see below) weight function defined on $[0,1]$ with $\int J(u)du = 1$. $\theta(F)$ is defined to be the solution of $\int J(u)Q(u-\theta(F))du = 0$ (where $Q = F^{-1}$) for all continuous F such that the integral is well-defined. For all F members of a specific translation family, $\theta(F)$ is a percentile of F. Which percentile $\theta(F)$ gives depends on the weight function and on the shape of F. For example, if J is symmetric about $1/2$ and the distribution determined by F is symmetric, $\theta(F)$ is the median of F. The weight function J can be used to construct the following process from which a slogcom estimator will be obtained.

Let $\{X_i, 1 \le i \le k\}_{(j)}$ denote the jth order statistic of the set $X_1,\ldots,X_k$. Then for s in $(0,1)$ fixed and ℓ positive, define

$Z_n(s+\ell/n) = \sum_{j=1}^{\ell} J(j/(\ell+1))\{X_{n\langle ns\rangle+i}, 1 \le i \le \ell\}_{(j)}/n$ and

$Z_n(s-\ell/n) = -\sum_{j=1}^{\ell} J(j/(\ell+1))\{X_{n\langle ns\rangle-i+1}, 1 \le i \le \ell\}_{(j)}/n$, where $\langle ns\rangle$ is the least integer greater than or equal to ns. Z_n can be thought of as the cumulative sum process of Corollaries 1 and 2 centered at s (i.e., $Z_n(t)-Z_n(s)$) with each sum of random variables replaced by the J-weighted sum of the order statistics of the same set of random variables. Extend Z_n to a continuous process on $[0,1]$ by linear interpolation and define $\mu_n(s) = \text{slogcom}(s)\ \{(t,Z_n(t)), 0 \le t \le 1\}$. For any finite set of integers A, define $N(A)$ to be the number of elements of A. Let $J_n(A)$ denote $\sum_{i\in A} J(j/(N(A)+1))\{X_{n,i}\ i \in A\}_{(j)}/n$. Then $\mu_n(s)$ can be written as $[N(L^*)\max J_n(L) + N(U^*)\min J_n(U)]/N(L^* \cup U^*)$, where the maximum is taken over the sets L of the form $\{i \le j \le \langle ns\rangle\}$ for some i, L^* is the largest such set for which the maximum is attained, the minimum is over sets U of the form $\{\langle ns\rangle + 1 \le j \le k\}$, and U^* is the largest such set for which the minimum is obtained. Note that L^* and U^* are disjoint.

Corollary 3.

If the following six assumptions are met, then

$$n^{\frac{N}{2N+1}}\left[\frac{(N+1)!}{\sigma^{2N}\mu^{(N)}(s)}\right]^{\frac{1}{2N+1}}(\mu_n(s)-\mu(s)) \xrightarrow[n\to\infty]{d} X^{((2N+1)^{-1})}\ .$$

1. $X_{ni} - \mu(\frac{i}{n})$ are independent, identically distributed random variables with cumulative distribution function F.

2. $\int J(u)Q(u)du = 0$, $\int J(u)du = 1$, and μ is non-decreasing on $(0,1)$.

3. J is continuously differentiable non-negative function whose J' satisfies a Hölder condition for some γ, $\frac{1}{2} < \gamma \leq 1$. The support of J is a compact subset of $(0,1)$.

4. $\sigma^2 = \iint J(F(x))J(F(y))F(\min(x,y))(1 - F(\max(x,y))dxdy > 0$.

5. μ has an Nth order derivative at s, $0 < s < 1$, where N is the smallest (finite) integer with $\mu^{(N)}(s) > 0$.

6. F is absolutely continuous, with strictly positive density f such that f converges to zero at infinity and f' is bounded.

The first two conditions describe the model and assert that the weight function J is appropriate. The third condition (which includes a requirement that J trim) is used to verify the Local Weak Convergence. The non-negativity of J will be used in the proof of the Hitting Time Condition. The fourth condition is more a definition than a condition, since the third condition ensures the integral is finite. The fifth condition describes the local behavior of μ at s. The sixth condition is a regularity condition used to obtain the Cornish-Fisher expansion needed to compute the drift component of the local weak convergence.

Corollary 3 shows that if the X_{ni} are all members of the translation family generated by the CDF F, the relative efficiency of two different isotonized linear combinations of order statistics with weight functions J_1 and J_2 is determined by the ratio $\sigma(J_1,F)/\sigma(J_2,F)$, where $\sigma(J,F) = \iint J(u)J(v)[\min(u,v)-u \cdot v]dQ(u)dQ(v)$. Corollary 1 shows that if F has finite variance, the same formula gives the efficiency of an isotonized linear combination of order statistics relative to the isotonized mean, although the weight function of the mean does not satisfy the conditions of Corollary 3. This same ratio is the asymptotic relative efficiency of two linear combinations of order statistics for estimating the location parameter of independent, identically distributed random variables whose distribution is a member of the location family generated by F. Therefore, all the comparisons known for a simple location problem carry over to isotonic estimation. In particular, if F does not have a variance, Corollary 3 applies, and the isotonized version of any linear combination of order statistics which trims will converge in the familiar manner. However, Corollary 1 does not apply. This extreme case shows that the isotonized mean is sensitive to wild observations and isotonized trimmed linear combinations of order statistics are more robust to heavy tails.

The proof of Corollary 3 can be described as showing that linear combinations of order statistics are similar enough to sums of independent random variables. For the details, see Leurgans (1978), Chapters IV and V.

The considerations of Example 2 imply that Corollary 3 can be extended to T_{ni} which are either other tractable deterministic sequences or order statistics from a suitable distribution.

Example 4: Estimation of a monotone density

Consider the estimation of a distribution F with support on the positive half-line using a sample of n observations. Grenander (1956) suggested an estimator for the case in which F has monotone (and hence decreasing) density. Grenander proved that this estimator is the restricted maximum likelihood nonparametric estimator of the density f. Grenander's estimator can be written as

$$f_n(x) = - \text{slogcom}\ (x)\{(t, - F_n(t)),\ t \geq 0\} \quad ,$$

where F_n is the empirical distribution function of the observations. Prakasa Rao (1969) proves the following result:

Corollary 4

Assume

1. f is a decreasing density on $[0,\infty)$.
2. $\{X_1,\ldots,X_n\}$ are a sample from this density.
3. f is differentiable at s and $f(s) < 0$.

Then

$$n^{1/3}\left[\frac{2}{f(s)f'(s)}\right]^{1/3}(f_n(s)-f(s)) \xrightarrow{d} X^{(1/3)} .$$

To obtain Prakasa Rao's result from Theorem 2.1, use $Z_n^{(t)} = F_n(t)$, $\mu(s) = -f(s)$, $\sigma(s) = f(s)$, $\rho(s) = f(s)/2$, and $p = 1/3$. The Hitting Time Condition can be obtained from Wald's Lemma. The local weak convergence condition involves $H_n(t)$, the empirical distribution function of $\{X_{ni} I_{\{X_{ni} \in (s, s+2cn^{-p})\}}, 1 \le i \le n\}$ evaluated at $s + 2cn^{-p}t$. $EH_n(t)$ is quite tractable and gives the centering needed. A routine finite dimensional distribution and tightness proof shows that $n(H_n(t) - EH_n(t))/(\sigma(s)(2cn^{1-p})^{1/2})$ converges weakly to $W(t)$.

Notice that Theorem 2.1 can be used to extend Corollary 5 in several directions. The density f may have first derivative zero. If some higher derivative of f is negative, rates and limiting distributions based on the order of this derivative are obtained. Furthermore, f need only be locally monotone in the sense of Example 1. Therefore f_n can be consistent even if the density function f is not totally monotone.

4. Proof of Theorem 2.1

Let Y_n denote the right-hand side of (1) and X_{nc} the same expression with μ_{nc} replacing μ_n. The theorem will be established in the following steps:

1. X_n for all c, $X_{nc} \xrightarrow[n\to\infty]{d} X_c$. ($X_c$ will be defined below)

2. $X_c \xrightarrow[c\to\infty]{d} X^{(p)}$.

3. $\lim_{c\to\infty} \underline{\lim}_{n\to\infty} P\{\mu_{nc}(s) = \mu_n(s)\} (= \lim_{c\to\infty} \underline{\lim}_{n\to\infty} P\{X_{nc} = Y_n\}) = 1.$

By Theorem 4.2 of Billingsley (1968) (p. 25), $Y_n \xrightarrow{d} X$, which implies Theorem 2.1.

Step 1. Since adding a line to a function increases the slope of the function's convex minorant by the slope of the line

$$\mu_{nc}(s) - \mu(s) = \text{slogcom}(s)\{(t,\ Z_n(t) - Z_n(s) - (t-s)\mu(s)),\ |t-s| \leq 2cn^{-p}\} \quad .$$

Translating the process so that the time scale of the function whose convex minorant is being obtained is $[-1, +1]$ and rescaling,

$$\frac{(2cn^{1-p})^{1/2}}{\sigma(s)} (\mu_{nc} - \mu(s)) = \text{slogcom}(0)$$

$$\left\{\left(t,\ \frac{n(Z_n(s+2cn^{-p}t) - Z_n(s)) - 2cn^{1-p}t\mu(s)}{(2cn^{1-p})^{1/2}\ \sigma(s)}\right),\ |t| \leq 1\right\} \quad .$$

Since slogcom(0) is a continuous functional on $C[-1,+1]$, the local weak convergence condition implies the expression above converges in distribution as $n \to \infty$ to $\text{slogcom}(0)\{(t, W(t) + (2c)^{1/(2p)}\rho(s)/\sigma(s)\ |t|^{(1+p)/2p}),\ |t| \leq 1\}$, which can be shown (using scale properties of the Wiener process) to equal $(2cK(s))^{1/2}\ X_c$, where $X_c = \text{slogcom}(0)\{(t, W(t) + |t|^{(1+p)/2p}),\ |t| \leq 2cK(s)\}$ and $K(s) = (\rho(s))^{2p}/(\sigma(s))^{2p}$. Dividing the last display by $(2cK(s))^{1/2}$, we see that for fixed c, X_{nc} converges in distribution to X_c.

Step 2. The only difference in the definitions of X_c and $X^{(p)}$ is that in X_c the set of points is restricted to $|t| \leq 2cK(s)$. Therefore to show X_c converges in distribution to X, it is necessary to show that large values of t do not affect the convex minorant of $W(t) + |t|^{(1+p)/2p}$. Since $p < 1$ implies the exponent of p is positive, the proof of this step follows from $W(t)/t \overset{a.s}{\to} 0$ $(t \to \infty)$, as is pointed out by Wright (1978). For an explicit proof in the case $p = 1/3$, see Prakasa Rao (1969) (Lemma 6.2, p. 34).

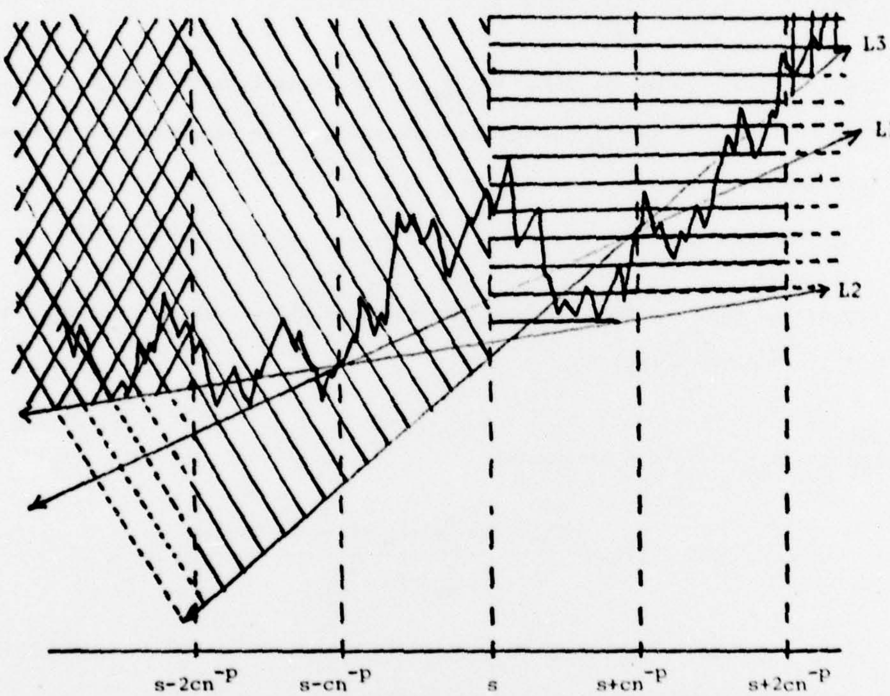

Figure 1.

Step 3. Figure 1 displays a realization of the process Z_n. At s, the greatest convex minorant of Z_n must lie entirely below L_1, the line connecting $Z_n(s + cn^{-p})$ and $Z_n(s - cn^{-p})$. Therefore no points of Z_n above this line can affect the slogcom(s){(t,Z_n(t)) $t \in \mathbb{R}$} and to establish this last step of the proof it suffices to show that $Z_n(t)$ lies above L1 for all t with $|t-s| > 2cn^{-p}$ with high probability. L1 has both random slope and random intercept. It is more convenient to work with L2 and L3, also superimposed on the diagram. L3 is the line through $(s + cn^{-p}, Z_n(s + cn^{-p}))$ with non-random slope $\bar{\mu}(n,c)$ and L2 is the line through $(s - cn^{-p}, Z_n(s - cn^{-p}))$ with slope $\mu(s) - (2cn^{-p})^{(1-p)/(2p)} \rho(s)$. It can be shown that if Z_n is above L_2 for $|t - (s - cn^{-p})| > cn^{-p}$ and above L_3 for $|t - (s + cn^{-p})| > cn^{-p}$ (as in the diagram), then Z_n lies above L_1 for $|t-s| > 2cn^{-p}$ and $Y_n = X_{nc}$. Therefore it suffices to show that the conditions of the probability that Z_n lies above two lines, for each of two separate intervals of t, is one in the appropriate limit. We shall show that the Hitting Time Conditions with $i = 1$ and the Local Weak Convergence Condition imply that Z_n lies above L3 for $t \leq s$ with appropriately high probability. The other three Hitting Time Conditions are used in the same manner, and then the Bonferroni Inequality can be used to complete the proof.

Thus it remains to show that $\lim_{c\to\infty} \underline{\lim}_{n\to\infty} p(n,c) = 1$, where (rearranging) $p(n,c) = P\{Z_n(s) - L3(s) \geq L3(t) - Z_n(t) + Z_n(s) - L3(s),\ t \leq s\}$. The probability that $Z_n(s) - L3(s)$ exceeds $L3(t) - Z_n(t) + Z_n(s) - L3(s)$ is less than the probability that $Z_n(s) - L3(s)$ is greater than a fixed constant $t(n,c)$ and that this same fixed constant is greater than $L3(t) - Z_n(t) + Z_n(s) - L3(s)$. Applying the Bonferroni Inequality to the intersection of the above two events, and recalling the definition of $L_1(t)$ in the Hitting Time Condition, it is easy to show that

$$(4) \qquad p(n,c) \geq P\{Z_n(s) - L3(s) \geq t(n,c)\} - P\{Z_n(t) - Z_n(s) < L_1(t),\ \text{some } t \leq s\} .$$

The Hitting Time Condition therefore implies the $\lim_{c\to\infty} \underline{\lim}_{n\to\infty}$ of the last term (minus sign included) is zero. Using the Local Weak Convergence Condition with $t = -1/2$ it can be shown (for details, see Leurgans (1978), chapter 3, section 3) that the $\underline{\lim}_{n\to\infty}$ of the first term is $1 - \Phi((\zeta-1)\lambda c^{(2p)^{-1}}\sqrt{2})$, where $0 < \zeta < 1$ (from the definition of $t(n,c)$), $\lambda = (2^{(1-p)/2p}-1)\rho(s)/(\sigma(s)\sqrt{2})$ is positive (because $p < 1$), and Φ is the cumulative distribution function of the standard normal distribution. Therefore the $\lim_{c\to\infty} \underline{\lim}_{n\to\infty}$ of the first term in (4) is 1, and the proof of the theorem is complete.

5. Remarks

Example 1 is a generalization of Brunk's Theorem 5.2. It should be remarked that Brunk's condition that "the observations satisfy Lindeberg's condition" can mislead the unwary: from the proof of example 1 we see that the observations must satisfy local Lindeberg conditions, which are unrelated to a global Lindeberg Condition. Wright's paper also generalizes Brunk's Theorem, and is the only paper known to the author with results for $N > 1$. Wright does not require that N be an integer and allows a different variance structure, but otherwise his results correspond to Example 1 and 2.

The estimators of Examples 3 and 4 have not been discussed in the literature. However, the results of Robertson and Wright (1975) include monotone estimators of the form $\bar{\mu}_n(s) = \max \min J_n(L \cup U)$, in the notation of Example 3. Robertson and Wright give conditions under which their minimax estimators are consistent for $\mu(s)$, but their methods do not give a rate of convergence. Corollary 3 gives such rates for slogcom estimators μ_n. It is natural to conjecture that μ_n and $\bar{\mu}_n$ have the same asymptotic behavior, even though μ_n and $\bar{\mu}_n$ are identical only in the case of Example 1. Unlike $\bar{\mu}_n(s)$, $\mu_n(s)$ is not always a monotone function of s. Isotonized percentiles of the Robertson and Wright type are also discussed by Casady and Cryer (1976).

Example 4 is due to Prakasa Rao (1969), and was the first asymptotic distribution obtained. Related hazard function estimators are discussed by Prakasa Rao (1970), Barlow and van Zwet (1970) and Barlow, et.al (1972). All of these results can be obtained from Theorem 2.1 and can be generalized in the manner described for Example 4.

Recall that the isotonized mean at s ($\hat{\mu}_n(s)$) is the mean of the X_{ni}'s over an adaptively chosen neighborhood of s. Theorem 5.8 of Barlow, et.al (1972) and Theorem 3.2 of Davis (1972) point out that for each s, if slightly wider deterministic windows centered at s are used, the resulting estimators converge more rapidly. However, this result appears to be the same sort of superefficiency result obtained in Example 1 for $N > 1$. In the case of Barlow, et. al, s must be at the center of every window. In Example 1, s must be exactly a point at which $\mu'(s) = 0$, but some other derivative is positive. If one is interested in estimation of an entire function, both kinds of s are isolated. Also, the deterministic window estimators need not give monotone estimators of $\mu(s)$.

The fact that μ_n can be consistent in some cases even when μ is not monotone is reminiscent of Theorem 3.4 of Barlow, et.al (1972), which states that likelihood ratio tests that some group means (normal errors, variances known and equal) exhibit a specified partial order against the null hypthesis that the means are all equal is an unbiased test of some alternatives which do not have the specified partial order against the same null hypothesis. The application to estimation does not appear to have been noted previously.

6. Acknowledgements

This paper represents an extension of parts of the author's Ph.D. thesis prepared in the Stanford Statistics Department, under the guidance of Thomas W. Sager. The dissertation was commenced while the author was a NIH Trainee and completed on an NSF Graduate Fellowship. David Siegmund suggested the use of the Dubins-Freedman and Hájek-Rényi Inequalities.

REFERENCES

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955), "An empirical distribution function for sampling with incomplete information," Ann. Math. Statist. 26: 641-647.

Barlow, R. E., Bartholomew, D. J., Bremmer, J. M., and Brunk, H. D. (1972), Statistical Inference under Order Restrictions, Wiley.

Barlow, R. E. and van Zwet, W. R. (1970), "Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: Strong consistency," (M. L. Puri, ed.) in Nonparametric Techniques in Statistical Inference, Cambridge: 159-173.

Billingsley, P. (1968), Convergence of Probability Measures, Wiley.

Brunk, H. D. (1970), "Estimation of isotonic regression," (M. L. Puri, ed.) in Nonparametric Techniques in Statistical Inference, Cambridge: 177-195.

Casady, R. J., and Cryer, J. D. (1976), "Monotone percentile regression," Ann. Stat. 4: 532-541.

Chernoff, H. (1964), "Estimation of the mode," Ann. Inst. Statist. Math. 16: 31-41.

Chow, Y. S., Robbins, H., and Siegmund, D. (1971), Great Expectations, Houghton Miflin Co.

Davis, H. T. (1972), "Nonparametric quantal response estimation procedures," Ph.D. thesis, Univ. of Chicago.

Dubins, L. E. and Freedman, D. A. (1965), "A sharper form of the Borel-Cantelli Lemma and the Strong Law,: Ann. Math. Statist. 36: 800-807.

Dubins, L. E. and Savage, L. J. (1965), "A Tchebycheff-like inequality for stochastic processes," Proc. Nat. Acad. Sci. 53: 274-275.

Grenander, U. (1956), "On the theory of mortality measurement. Part II," Skand. Akt., 39: 125-153.

Leurgans, S. (1978), "Asymptotic distribution theory in generalized isotonic regression," Tech. Report, Dept. of Statistics, Stanford University.

Parson, L. V. (1975), "Distribution theory of isotonic estimators," Ph.D. thesis, Univ. of Iowa.

Prakasa Rao, B. L. S. (1966), "Asymptotic distributions in some non-regular statistical problems," Tech. Report #9, Dept. of Statistics and Probability, Michigan State Univ.

Prakasa Rao, B. L. S. (1969), "Estimation of a unimodal density," Sankhya (A) 31: 23-36.

Prakasa Rao, B. L. S. (1970), "Estimation for distributions with monotone failure rate," Ann. Math. Statist. 41: 507-519.

Robertson, T. and Wright, F. T. (1975), "Consistency in generalized isotonic regression," Ann. Stat. 3: 350-362.

Van Eeden, C. (1956), "Maximum likelihood estimation of ordered probablities," Proc. K. ned. Akad. Wet (A) 59/ Indag. Math. 18: 444-455.

Wright, F. T. (1978), "The Asymptotic Behavior of Monotone Regression Estimates," submitted for publication.

SL/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
#1946

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

(9) Technical

4. TITLE (and Subtitle)
(6) ASYMPTOTIC DISTRIBUTIONS OF SLOPE OF GREATEST CONVEX MINORANT ESTIMATORS.

5. TYPE OF REPORT & PERIOD COVERED
Summary Report, ~~no specific~~ [redacted]

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
(10) Sue/Leurgans

8. CONTRACT OR GRANT NUMBER(s)
(15) DAAG29-75-C-0024
✓NSF-MCS77-16974, [redacted]

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Work Unit Number 4 - Probability, Statistics, and Combinatorics

11. CONTROLLING OFFICE NAME AND ADDRESS
See Item 18 below

12. REPORT DATE
(11) Apr[redacted]979

13. NUMBER OF PAGES
22

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
(12) 27p.

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

(14) MRC-TSR-1946

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, D. C. 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Isotonic estimation, asymptotic distribution theory.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Isotonic estimation involves the estimator of a function which is known to be increasing with respect to a specified partial order. For the case of a linear order, a general theorem is given which simplifies and extends the techniques of Prakasa Rao (1966) and Brunk (1970). Sufficient conditions for a specified limit distribution to obtain are expressed in terms of a local condition and a global condition. The theorem is applied to several examples. The first example is estimation of a monotone function μ on $(0,1)$ based on

next page

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

221 200

ABSTRACT (continued)

sub ni mu

observations $(i/n, X_{ni})$, where $EX_{ni} = \mu(i/n)$. In the second example, i/n is replaced by random T_{ni}. Robust estimators for this problem are described. Estimation of a monotone density function is also discussed. It is shown that the rate of convergence depends on the order of first non-zero derivative and that this result can obtain even if the function is not monotone over its entire domain.